SEQUENTIAL DIAGNOSIS BY COMPUTER

299-67

G. Anthony Gorry*

and

G. Octo Barnett**

SEQUENTIAL DIAGNOSIS BY COMPUTER

299-67

G. Anthony Gorry*

and

G. Octo Barnett**

*Assistant Professor, Sloan School of Management, M.I.T.

**Director, Laboratory of Computer Science, Massachusetts General Hospital

Work reported herein was supported (in part) by NIH grant GM-15287-03 and (in part) by Project MAC, an M.I.T. research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract Number Nonr-4102(01).

SEQUENTIAL DIAGNOSIS BY COMPUTER

1. Introduction

The use of digital computers in diagnosis has received increasing attention in recent years. Much of this interest derives from the general desire to provide tools and techniques to facilitate the solution of diagnostic problems by the physician. Because a significant part of diagnosis consists of sifting out numerous possibilities and developing hypotheses about the state of health of the patient, there has been considerable interest in the application of automatic data processing in this activity. The ability of a computer to store extremely large amounts of data, to enumerate many possibilities, and to perform complex logical operations suggest its potential value in diagnosis. These capabilities alone, however, are not sufficient for the purpose. Before a computer can be used to significant advantage in diagnosis, precise procedures must be formulated for the means of inference required to deduce the clinical state of the patient from observed signs and symptoms.

As the latter problem is examined more closely, two general approaches become apparent. First one can attempt to capture the processes used by physicians in a manner sufficiently precise to permit their use in a computer program. There are several drawbacks to this approach. The most

obvious is the current lack of understanding of the processes employed by doctors. The manner in which a physician interprets clues from the patient may be extremely difficult to describe with the required precision. Even if such were not the case, however, the simulation of human diagnostic processes on a computer may fail to exploit the comparative advantages of the machine. It seems quite unlikely that the diagnostic processes of doctors would prove to be those most suited for a computer. On the other hand, diagnosis as performed by humans may be a source of valuable insights into those capabilities which must be developed for the computer. Because of these difficulties, most of the studies of computer-aided diagnosis have focused on the development of strategies particularly suited for a computer, without the restriction that they also be suited for humans.

In the typical situation, a doctor confronted with a particular diagnostic problem must interpret the available evidence (observed signs, symptoms, etc.) in terms of his medical knowledge and experience. In other words, he employs a method of deduction which can accommodate both his understanding of diseases in general, and the individual instance represented by the patient before him. In this discussion, the processes involved in the interpretation of this clinical situation in terms of the total medical experience will be termed the <u>inference function</u> of diagnosis.

A number of studies of computer-aided diagnosis have focused on the use of Bayes rule as the inference function for the machine. Because Bayes rule has been discussed in a number of papers (1, 2), it will be reviewed only in general terms here.

Bayes rule provides a means for computing the probability that a patient has a particular disease given an observed set of signs and symptoms. This computation requires the *a priori* probability of each disease in question and the conditional probabilities of sign and symptom patterns given each disease. These probabilities can be obtained from an analysis of historical data. Generally the analysis is facilitated by some simplifying assumptions about independence of various signs and symptoms. The outcome of this analysis, then, is some (perhaps large) number of probabilities which are used by the computer in diagnosis. Through the use of Bayes rule, the inference function is transformed from the qualitative and individualistic type of process often attributed to doctors into a simple computational process eminently suited for a computer.

Some studies of Bayes rule have suggested that it may have considerable value in computer-aided diagnosis. In spite of the simplifications introduced to develop an inference function appropriate for a computer, the results obtained with such a function are quite good. These studies of Bayes rule in diagnosis have followed basically the same pattern. A particular medical area was analysed to obtain the requisite probabilities. If insufficient data were available for the determination of a particular probability, the investigators made an estimate of it which reflected their experience in the area. The probabilities obtained from this analysis constituted the 'experience' which a program based on Bayes rule could use in diagnosis. For each patient, a given set of tests were performed, and the results of these tests were input to the program. The program computed

the probability distribution for the diseases in question through the use of Bayes rule. This distribution constituted the diagnosis arrived at by the program.

Studies of this general nature have been conducted by Warner and his associates in the field of congenital heart disease (2, 3); Lodwick (4), in primary bone tumors; and Overall and Williams (5), in thyroid function. In each case, the investigators reported that a program based on Bayes rule performed at a level approaching that of an expert in the respective field. The success of these efforts in diverse medical areas has strengthened the belief that Bayes rule may be of considerable value in computer-aided diagnosis.

There is, however, an important shortcoming inherent in the approach used in these studies. The programs are designed to transform a set of test results into a diagnosis in one stage. It is assumed that all the relevant signs and symptoms are known <u>before</u> the program is invoked, and the only task of the program is to develop a diagnosis from this information. While this may be an appropriate strategy in the particular areas cited, it is not appropriate for more general use. As evidence of this inadequacy, consider again the manner in which a doctor performs diagnosis.

In general, a doctor does not have sufficient information initially on which to base a diagnosis. The information he does have and his medical knowledge and experience permit him to develop a <u>current view</u> of the problem. This view exerts a considerable effect on the strategies which he will employ later in the diagnostic process. The doctor seeks additional information

about the disease state of the patient in order to improve this current view. He has a variety of means to obtain more information such as questions and laboratory tests. (For convenience, the term test will be used here to denote all those questions, laboratory procedures, etc., which can yield information relevant to a diagnosis.) He employs those tests which he expects to yield results of significant value in improving his current view of the problem. Note that the doctor's assessment of the potential value of a test will be strongly influenced by his current view. For example, whether a chest x-ray will be taken may be determined in large part by whether the doctor believes tuberculosis to be the disease present.

The value of information obtained from a test is determined by the contribution this information makes to reducing the likelihood of a misdiagnosis and its associated cost. The more information the doctor obtains about the condition of the patient, in general, the less the risk of a possible misdiagnosis will be. Hence the doctor is inclined to perform many tests. On the other hand, these tests are not without some costs in terms of patient discomfort, time of skilled persons, money, etc. Thus there is a conflicting tendency to hold the number of diagnostic tests to a minimum.

The doctor resolves these conflicting tendencies by performing sequential diagnosis. Based on his current view of the problem, he evaluates the choices available to him. He can cease testing and make a final diagnosis. In doing so, he incurs the risks associated with possible

misdiagnoses. On the other hand, he can select a test to perform in the hopes of gaining additional information upon which to base his diagnosis. In this case, he incurs the cost of the test selected. When the results of the test become available and are incorporated into his view of the problem, he is again faced with this same decision problem. Thus, in reaching a diagnosis, a physician solves a sequence of similar decision problems. At each stage in the diagnosis, he balances the costs of further testing against the expected reduction in the risk of misdiagnosis which the test results will permit. Tests are performed as long as they are expected to reduce the risk of misdiagnosis by more than their cost. When, in the opinion of the physician, no test possesses this property, he will make his final diagnosis. Because he repetitively updates his current view of the problem and re-evaluates the potential usefulness of tests, the physician is able to develop relatively efficient testing sequences in diagnosis.

Although this description of the diagnostic process employed by doctors is undoubtedly simplified, it does emphasize the fundamental role that sequential decision-making plays in the process. It seems clear that it will be necessary for a computer to exploit a similar capability in diagnosis. The studies of the application of computers to medical diagnosis cited above were concerned with the <u>inference function</u> of diagnosis. They did not deal with the equally important process of selecting diagnostic tests. This process is referred to here as the <u>test selection function</u>.

The remainder of this paper is devoted to a general review of some research on the problem of sequential diagnosis by computer. The emphasis is on the concepts involved, and the discussion is non-technical. Readers interested in more detail are referred to other accounts of this work (1, 6, 7).

2. A Computer Program for Sequential Diagnosis

A computer program for sequential diagnosis has been designed and is currently in use on the time-sharing computer system at Project MAC. This system simultaneously serves many users at remote terminals (typewriter-like devices). It is possible for a user to interact in an almost instantaneous fashion with a program in the system, engaging in a dialogue with that program. The program discussed here exploits this type of interaction with the user.

There are three major parts of the program: the information structure, the inference function, and the test selection function. These functions and their interrelations are discussed in this section.

The _information structure_ constitutes the medical experience available to the program. By changing the information structure, one can convert the program for use in a new problem area. The inference function and the test selection function are constant for all applications. The information in the information structure is of two basic kinds: probabilities and costs. The inference function employs Bayes rule and hence requires _a priori_

probabilities for the diseases in question and conditional probabilities for various signs and symptoms given these diseases. As noted above, these probabilities are obtained from an analysis of historical data and the opinions of experts in the given field. Thus the program "understands" diseases only in terms of signs and symptoms and the associated probabilities.

The information structure also includes the costs of the tests which the program can select. As was mentioned previously, the term test refers to an activity which yields information about the presence or absence of a given sign or symptom or laboratory observation. There is at least one test for each sign or symptom or laboratory result which may be relevant in diagnosis, although a given test may suffice for a number of signs, symptoms, and laboratory results. Associated with each test in the information structure is a cost of using this test. Tests may differ markedly in cost, and the program must take cognizance of this in developing a testing strategy. Also included in the information structure are the costs associated with possible misdiagnoses. The particular form of these costs is a number for each ordered pair of diseases which is the cost of misdiagnosing one disease as the other. For example, the cost of misdiagnosing a malignant tumor as benign might be 1,000,000, while the cost for the opposite misdiagnosis might be 1,000. The importance of these costs will be discussed below.

In order to employ the program in diagnosis in a given problem area, one first prepares an information structure for the area. Once such a structure has been made available to the program, it can be used to solve

diagnostic problems. The basic mode of operation of the program is as follows. From his terminal, the user informs the program of the initial signs and symptoms associated with a particular diagnostic problem. The program employs the inference function to obtain a current view of the problem based on these clues and the experience available in the information structure. The test selection function is then invoked to select a test for the user to perform. The test is performed by the user, and when the results are reported to the program, the inference function is again employed to update the current view. Now another test is selected by the test selection function in accordance with this newly updated view of the problem. This iterative process continues until the test selection function determines that testing should be discontinued and a final diagnosis is made. At each stage of this process, the user is aware of the program's view of the problem, and he can follow its development to a final diagnosis.

The inference function uses Bayes rule to update a probability distribution which is the current view taken by the program. This updating takes place after each new test result is obtained. The updated probability distribution is one of the factors considered by the test selection function in determining the strategy which the diagnostic program is to follow.

The problem for the test selection function is to select one of a number of decision alternatives for the diagnostic program. There is one of these decision alternatives corresponding to the selection of each of the potentially useful tests and one corresponding to the cessation of testing

(making a final diagnosis). In assessing the potential value of each of these alternatives, the test selection function draws in information from several sources.

By considering the probability distribution associated with the current view of the problem and the costs of possible misdiagnoses, the test selection function determines the best diagnosis to make _assuming_ that no further tests are to be performed. This diagnosis is the one which minimizes the expected cost of a misdiagnosis, and this cost is taken by the test selection function as the cost of the final diagnosis alternative.

In evaluating the usefulness of a particular test, the function considers the current view of the problem (as developed by the inference function), the cost of the test, and the likelihood of possible test results. For a possible test result, the inference function can be used to _simulate_ the change in the current view which would occur if the result were obtained, and in turn, the expected cost of a final diagnosis based on this new view. Thus, for a specific result of a given test, the reduction in the expected cost of a final diagnosis can be determined. The sum of the expected reduction in the cost of misdiagnosis associated with each test result _weighted_ by the probability that that result will be obtained is the expected reduction in the cost of misdiagnosis for the given test. This weighting is used because even though a specific test result may significantly reduce the expected cost of misdiagnosis, it may be so improbable (based on the current view and the probabilities in the information structure) that it contributes little to the total expected value of

performing a test. This weighted sum can be compared to the cost of the test in assessing its potential usefulness. This type of analysis can be extended to sequences of tests at the expense of increased computation.

By analysing the available tests in this manner, the test selection function can determine the best test to perform at this stage in the diagnosis. A final diagnosis is indicated whenever no test promises to reduce the expected cost of a final diagnosis by more than the cost of using the test. Because the diagnostic program operates in a sequential mode, it can continually re-evaluate the potential usefulness of available tests in keeping with changes in its current view of the problem. In this way it is able to employ reasonably efficient testing strategies which properly reflect the various costs associated with the diagnostic problem.

3. Results of Experimentation with Sequential Diagnosis

To date, the diagnostic program outlined above has been employed in two distinct medical areas, the diagnosis of primary bone tumors, and the diagnosis of congenital heart disease. Some results obtained from the latter application are presented here in support of observations on the value of sequential decision-making in computer-aided diagnosis. From data supplied by Dr. Warner and his associates, an information structure for the congenital heart disease area was constructed for use with the diagnostic program. There were thirty-five diseases in question and a total of fifty-two signs and symptoms were available for use in diagnosis. The

latter consisted of x-ray findings, murmers, physical signs, etc. Because Warner has reviewed the definition of the problem (2), only general observations about it will be offered here.

Thirty-four tests were defined for use by the program. These tests included, for example, a test which revealed the age of the patient and one which revealed the presence and extent of cyanosis. These tests were used by the program to elicit information about the cases with which it was presented. Warner also supplied several hundred case histories with definitive diagnoses. These cases were used to evaluate the effectiveness of the sequential approach in computer-aided diagnosis. The procedure for using the program in these tests was as follows. The results of three tests (always the same three) were given to the program as an initial definition of a diagnostic problem. The three tests chosen concerned the age of the patient and the presence of cyanosis and the squatting symptom because these results were considered basic information and easily obtainable. The program then began to select new tests to obtain additional information. The response to each test by the user was in keeping with the information in the case history. In this experiment, all diseases were considered equally serious, and the cost of each test was taken to be unity. The final diagnosis reached by the program was recorded, as well as the tests required to reach this diagnosis. Then the results of tests <u>not</u> chosen by the program were revealed to the inference function, so that the program could obtain a <u>complete</u> <u>diagnosis</u> (i.e., a diagnosis based on <u>all</u> the available signs and symptoms). For a given case history, both the sequential

and the complete diagnoses were compared to the actual diagnosis provided with the case. Because Warner and his co-workers have shown that a complete diagnosis is comparable on the average to that obtained by experts in the field (2), this diagnosis provided a standard of comparison for the sequential diagnosis.

The results of this experiment are presented in detail in another paper (7); only a few summary figures will be included here. For convenience, a performance measure suggested by Warner (4) was used. This measure is the product of the average probability assigned to the actual disease in the final diagnosis and the fraction of the cases in which the actual disease was assigned a probability greater than 0.01. The motivation for this measure is that it is sensitive to both the average probability of the actual disease and the fraction of "complete misses" by the program. The value of this performance measure obtained in this experiment varied as a function of the particular disease. In general, the more common the disease, the better the performance as reflected by this measure. Because the set of cases provided by Warner did not correspond exactly to the expected frequency of various disease types, a normalized measure was computed. For the sequential diagnosis, this normalized measure was 0.57, and for the complete diagnosis, it was 0.59.

In terms of this measure, then, there was essentially no difference between the diagnostic accuracy achieved by the sequential and the complete modes of diagnosis. When the number of tests required by these two modes is examined, however, a striking difference is observed. As was noted,

a total of thirty-one tests are required for the program to discover all of the relevant signs and symptoms. The average number of tests required by the sequential version of the program, however, was only 6.9. This striking reduction in the number of diagnostic tests employed results from the use of a test selection function which continually assesses the potential usefulness of the available tests. In this problem area, a program possessing such a capability requires only a few well-chosen tests to achieve the accuracy of complete diagnosis (which in turn is comparable to that attained by expert human diagnosticians). Because any measure of overall diagnostic performance should include both accuracy and testing cost, these results suggest quite strongly the value of this sequential approach for computer-aided diagnosis.

4. Discussion

A realistic appraisal of the usefulness of the sequential diagnosis approach will require a much more extensive evaluation than that provided in this preliminary experimentation. Only two hundred and fifty cases were studied, and all these cases were from the same general area. Although the more limited studies of the bone tumor problem support the comments above, more problem areas need to be explored with the program. In addition there are many specific problems associated with computer-aided diagnosis which require more attention. Nonetheless a program which includes both an inference function and a test selection function has considerable appeal as a means for utilizing the computer in diagnostic problem-solving because it is focused on two fundamental aspects of the problem.

While there are a number of unresolved problems, one requirement for such a program merits special attention. In the description of the information structure employed by the program, the concepts of the costs of tests and the costs of misdiagnoses were introduced. These costs play a fundamental role in the determination of testing strategies. If, for example, the average cost of testing is small relative to that of possible misdiagnoses, the average number of tests used by the program will increase. This of course is a desirable effect. It does mean, however, that a reasonably accurate assessment of the relative magnitudes of the various costs involved may be required to obtain the best performance from the program. This need becomes even more pronounced when one considers that the selection of a final diagnosis is also affected by costs. That is, decisions about the disease of the patient reflect not only the likelihoods of possible diseases, but the potential seriousness of misdiagnosing these diseases.

The most significant problem resulting from uncertainty about the various costs discussed above is that it is very difficult to properly measure diagnostic performance. As noted, any measure of diagnostic performance should be based on <u>total cost</u>, both the cost of testing, and the cost of misdiagnosis. To ignore either of these costs in measuring performance is to seriously distort the true purpose of diagnosis.

The determination of these costs (on some common scale) undoubtedly is an extremely difficult task in any actual problem area. One possibility is to study the sensitivity of diagnostic performance to variations in these costs, and concentrate on determining good estimates of these.

In spite of the significant problems which remain to be solved, the initial experience with a program based on a sequential view of diagnosis has been very encouraging. It is believed that the advantage of such a program will be even more pronounced in areas where groups of diseases require distinct sets of tests for identification. Here the reduction in the number of tests required by the sequential program as opposed to one using the complete set of tests will be even more striking. In regard to the general problem area in which a great number and variety of diagnoses must be considered, the sequential approach is the only feasible one for computer-aided diagnosis.

REFERENCES

(1) Gorry, G. A., and G. O. Barnett, "Experience with a Model of Sequential Diagnosis", Computers in Biomedical Research, Vol. 1, No. 5 (to appear).

(2) Warner, H. R., A. F. Toronto, and L. G. Veasy, "Experience with Baye's Theorem for Computer Diagnosis of Congenital Heart Disease", Annals of New York Academy of Sciences 115, 2, 1964.

(3) Toronto, A. F., L. G. Veasy, and H. R. Warner, "Evaluation of a Computer Program for Diagnosis of Congenital Heart Disease", Progress in Cardiovascular Diseases, 5, 362, 1963.

(4) Lodwick, G. S., et al., "Computer Diagnosis of Primary Bone Tumors: A Preliminary Report", Radiology, 80, 1963.

(5) Overall, J. E., and C. M. Williams, "Conditional Probability Program for Diagnosis of Thyroid Function", Journal of the American Medical Association, 183, 5.

(6) Gorry, G. A., "A System for Computer-Aided Diagnosis", Project MAC Technical Report MAC-44 (THESIS).

(7) ____________, "Strategies for Computer-Aided Diagnosis", Mathematical Biosciences (to appear).

Date Due

JAN 27	JUN 14 2000	
DEC 22 '81		
JUN 3 '83		
MAY 1 '84		
MAY 2 2 1987		
MAR 08 1990		